

De la rigidez de la imprenta a la flexibilidad digital: hacia un esqueleto interno en la gestión documental

Karim Gherab Martín, INDRA

Introducción

La tecnología digital ofrece a los bibliotecarios la posibilidad de pensar en modelos alternativos de gestión documental, la oportunidad de abandonar la rigidez del soporte impreso para abrazar nuevas formas de organizar los objetos digitales, construyendo un esqueleto flexible y obediente a las infinitas posibilidades que se le ofrezcan. Ello es posible adoptando un nuevo modelo de *descriptores internos* que, sin renunciar a los tradicionales *descriptores externos*, permita hacer mayor énfasis en las relaciones existentes entre documentos (*relata*) y en las plurales opiniones que los lectores tengan a bien aportar. Dicho modelo configura una estructura interior que ofrece innumerables itinerarios para recuperar un determinado documento, complementando la habitual y precisa recuperación del documento mediante la catalogación externa con una búsqueda más espontánea e indirecta que facilite los saltos interdisciplinarios.

Un caso particular de lo dicho es el caso de la *indexación por citas* (referencias bibliográficas, índices onomásticos, etc.) en contraposición con la tradicional *indexación por temas*. El modelo o topología general para organizar de este modo el conocimiento podría denominarse como *modelo popperiano*.

Esta propuesta general ha tenido sus realizaciones particulares en ejemplos prácticos exitosos que se han ido aplicando en diversos sectores de registro del saber, como son:

- La jurisprudencia en el Derecho anglosajón (el caso de la firma consultora *Shepard Citations*).
- Los índices de impacto aplicados por Eugene Garfield (a partir de 1955) en los modelos de gestión de patentes y en la ciencia contemporánea, en este caso a través del *Science Citation Index* publicado anualmente por el Institute of Scientific Information (ISI) con sede en Philadelphia.
- La Enciclopedia Británica.
- Google y su técnica *PageRank*, patentada en 2001 por la Universidad de Stanford y que se basa en el *Science Citation Index* mencionado arriba.

El caso Shepard Citations

Es ciertamente curioso que la forma internalista (descriptores internos) de plasmar el conocimiento por citas no haya tenido su inspiración en la manera de proceder de las ciencias llamadas *duras*, sino que provenga de una de las ciencias *blandas*, del Derecho.

La idea de Eugene Garfield de aplicar la indexación por citas a las patentes y, posteriormente, a la ciencia se basó en el modo de proceder normal en los tribunales estadounidenses. En efecto, los juristas norteamericanos citan casos anteriores porque sientan precedente, y tal recurso a un caso anterior de sentencia favorable puede servir de poderoso argumento para quien lo trae al caso, de modo que hacen referencia al volumen y a la página en el que dicho caso se encuentra registrado o bien al artículo, capítulo, sección y publicación en el que se encuentra determinado estatuto. Así, nació una publicación en el último cuarto del siglo XIX, *Shepard's Citations*, perteneciente a la empresa consultora Franck Shepard Company, que se encargaba de proveer a los juristas de tales citas (casos, estatutos y sentencias) y, especialmente, de mostrar la *historia* de un caso registrando todos los casos posteriores que lo han citado, informando así del resultado de la sentencia de los casos posteriores que citaron ese caso. La indexación de los casos jurídicos detallaba incluso, mediante una sencilla codificación o etiquetaje por letras el historial de cada caso, indicando con la letra “e” aquellos casos en los que el tribunal dio una explicación del caso original citado, con la letra “a” para denotar que en esa ocasión el tribunal confirmó la validez del precedente, con la letra “d” para especificar que se introdujeron una serie de aclaraciones o distinciones respecto del caso original, etc. Veamos un ejemplo (Adair, 1955):

	101 Mass 210
	112 Mass 65
e	130 Mass 89
	165 Mass 210
d	192 Mass 69
	205 Mass 113
e	212 Mass 173
	221 Mass 210
a	281 U.S. 63
	35 H.L.R. 76

El caso original a partir del cual el abogado comienza su investigación es el caso 101 que aparece en la página 210 del apartado de informes correspondientes al estado de Massachussets (EEUU). Los casos que aparecen en el listado anterior son todos aquellos casos que citaron el caso original, de modo que en el caso 130 Mass 89 el tribunal dio una explicación del caso original (denotándolo con el código “e”), en el caso 192 Mass 69 el tribunal introdujo una aclaración o distinción (“d”) con respecto del caso original limitando el área de validez del caso como precedente, y en el caso 281 U.S. 63 el Tribunal Supremo de los EEUU aseveró (“a”) que el caso representaba un buen precedente como ley, lo cual fue publicado en el *Harvard Law Review* (H.L.R.).

Un listado completo de las etiquetas utilizadas así como la reproducción de un pedazo de página de dicho catálogo se muestran en la imagen siguiente:

Shepard Abbreviations		CITATIONS TO LAW, AS PUBLISHED IN COURT DECISIONS									
History of Case	Treatment of Case	CITED COURT DECISIONS									
a (affirmed)	c (criticised)	Vol. 386	- 171 -	- 714 -	- 780 -	- 167 -	- 541 -	- 175 -	272So2d624	16MdA529	
cc (connected case)	d (distinguished)	352FS1411	352FS1411	351FS588	471F2d1091	471F2d190	471F2d1091	31NY951	Calif	292Min402	
D (dismissed)	e (explained)	So C	So C	h351FS1955	127CaA515	352FS708	127CaA515		105CaR677	96PRR389	
m (modified)	f (followed)	194S2403	194S2403	57FRD199	521I2d480		521I2d480		105CaR719	96PRR392	
r (reversed)	h (harmonized)	- 213 -	- 213 -	- 724 -	- 213 -	- 213 -	- 213 -	- 218 -	470F2d153	106CaR107	96PRR393
s (same case)	j (dissenting opinion)	352FS1741	352FS1741	352FS1001	470F2d1695	470F2d1695	194S2319		470F2d162	106CaR110	96PRR394
S (superseded)	L (limited)	174Col271	174Col271		NY		NY		470F2d380	106CaR112	96PRR711
v (vacated)	o (overruled)	174Col537	174Col537		- 244 -	- 244 -	340S2d612		470F2d380	505P2d531	108RI531
	p (parallel)	71A109	71A109		- 1 -	- 1 -			470F2d718	505P2d534	1TnCr123
	q (questioned)	16MdA	16MdA		1F2d1578	1F2d1578			d470F2d949	505P2d536	1TnCr185
		72Msc2	72Msc2		1FS168	1FS168			d471F2d119	505P2d536	1TnCr350
		521I2d480	521I2d480		470F2d119	470F2d119			471F2d425	505P2d3	1TnCr417
					470F2d119	470F2d119					1TnCr692

El caso de la ciencia

Garfield vio rápidamente la validez de este esquema para la indexación del conocimiento en la ciencia pero probó primero con los fondos bibliográficos de patentes en el área química (Garfield, 1957), la disciplina de la que él mismo provenía. En efecto, el funcionamiento de las oficinas de patentes es similar al esquema mostrado puesto que los examinadores han de asegurarse que el invento es realmente original y recurren a los fondos bibliográficos para confirmarlo. Así, deben respaldar sus conclusiones haciendo referencia a estudios y patentes anteriores de modo que todo el historial de una rama de invención quede registrado con precisión y sea recuperado eficazmente y sin lagunas. Naturalmente, en cualquiera de los tres casos (Derecho, Patentes y Ciencia) es preciso iniciar las investigaciones a partir de un caso, una patente o un artículo concreto, lo cual puede no ser sencillo si no se conoce el área de conocimiento correspondiente. Sin embargo, la dificultad es superable en cuanto alguien nos indique un artículo por el que comenzar a investigar, o bien observando con detenimiento las referencias de unos cuantos artículos relevantes al caso, por ejemplo, mirando un índice temático. Vemos pues que, lejos de abogar por la sustitución de la indexación por temas por la indexación por citas, ésta complementa a aquella, y a la inversa.

En 1964, Garfield (1964) presentó el *Science Citation Index* (SCI), catálogo de indexación por citas para los artículos científicos, un indicador que, en muy poco tiempo, revolucionaría no solo la sociología de la ciencia sino también las políticas de ciencia y tecnología. Garfield consideraba factible analizar las mil revistas científicas más prestigiosas y estudiar para cada disciplina la dinámica de citas. Más tarde, se publicarían por parte del Institute of Scientific Information (ISI), con sede en Filadelfia, fundado y dirigido por Eugene Garfield, el *Social Sciences Citation Index* (SSCI) que recogería la dinámica de citas de las revistas correspondientes a las ciencias sociales, economía, sociología, política, etc., el *Arts & Humanities Citation Index* (AHCI) para las artes y las humanidades y, finalmente, el *Journal Citation Report* (JCR) cuyo objetivo sería dar un factor de impacto (*Impact Factor*) de las revistas científicas más prestigiosas, en cualquier área.

Una propiedad importante de la indexación por citas presente en la ciencia y que también es visible en el caso *Shepard Citations* es la estructura espacio-temporal que tales modelos resaltan. Un artículo citado es un artículo anterior en el tiempo en tanto que si nos movemos por las ramas del árbol de las citas, los artículos citantes que aparecen en los catálogos de los ejemplos anteriores son más recientes. El tiempo mismo se convierte así en un descriptor interno, un descriptor que no es explícitamente

mencionado en ningún sitio pero que es direccionalmente visible. Tratando ahora la “dimensión espacial”, la localización de un artículo no es absoluta en el sentido que Newton asignaba al espacio absoluto, sino que se asemeja más a la lógica de ordenación espacial leibniziana, que se inclinaba por definir un lugar espacial en función de su relación topológica con los demás. Este sello de lugar y tiempo es el que permite conocer la “distancia leibniziana” de un artículo a cualquiera de las disciplinas a las que podría legítimamente pertenecer.

Enciclopedia Británica

Sin entrar en detalle en la manera en que está organizada la Enciclopedia Británica, no está de más mencionarla dado que su metodología se asemeja a los modelos anteriormente descritos. Realmente en este caso no hay necesidad de valorar unos contenidos sobre otros, ya que sus definiciones y demás contenidos han sido redactados por expertos en cada materia. Sin embargo, se echa en falta en ella una mayor claridad en cuanto a la relación temporal de sus contenidos, carencia que se ha suplido con la publicación de ediciones cada cierto tiempo. Pero no es la Enciclopedia Británica la culpable de esta limitación, son la propia tecnología en la que se ha sustentado hasta hace bien poco: la imprenta. Las ligaduras impuestas por esta técnica varias veces centenaria y a la que no queremos o no podemos renunciar tan fácilmente, no dejan ver con claridad la inmensa flexibilidad que nos ofrecen los procedimientos digitales. Las técnicas de indexación digital nos permitirán ir más allá de la topología de la *Británica*, de tal manera que no nos conformaremos con una topología de red, una imagen que todo el mundo usa hoy como metáfora de organización de la información, sino que emergerá una *selva*, esto es, un número incontable de árboles interconectados donde las ramas apuntan al futuro (los artículos citantes) y las raíces se sostienen en el pasado (artículos citados).

Obviamente, una selva como la esbozada es muy diferente a aquella Enciclopedia Francesa que lideraron Diderot y D’Alembert, en el que los distintos saberes se bifurcaban en ramas a partir de un tronco común, separándose ya para siempre. El modelo de las tres ramas primitivas (memoria, razón, imaginación), basado en las propuestas de Avicena y F. Bacon, no ha perdido validez a día de hoy, pero requiere ser repensado en función de la nueva tecnología con la que estamos transformando el mundo. Tenemos que volver a reflexionar acerca de las relaciones que *memoria*, *razón* e *imaginación* tienen entre sí en el nuevo entorno digital.

La tecnología *PageRank* de Google

Muchas son las razones del éxito de Google y qué duda cabe que su modelo de negocio es ciertamente innovador. Sin embargo, la mayor innovación (que no invención) de sus jóvenes fundadores, Larry Page y Sergey Brin, ha sido la de aplicar a Internet el método que Garfield había introducido en la indexación de artículos científicos. Page y Brin no leyeron directamente a Garfield pero vieron rápidamente su utilidad al leer un artículo sobre los indicadores (*factores de impacto*) utilizados por los bibliotecarios al valorar artículos y revistas en el ámbito académico. Sin entrar en disquisiciones acerca de algunos efectos perniciosos del *factor de impacto*, lo cierto es que Google destacó sobre el resto de buscadores, que por aquel entonces más que buscadores eran portales que

presentaban árboles de navegación temática elaborados por editores de los portales mismos.

Yahoo! ofrecía este modelo a sus usuarios, un detalle que todos agradecíamos cuando aún no nos manejábamos con soltura por la Red y que agradecían incluso los propios editores del portal cuando la información disponible era todavía escasa. Obviamente, el crecimiento exponencial o geométrico de páginas en Internet ha rebasado cualquier capacidad de crecimiento aritmético en el esfuerzo de los editores, de manera que se intuía que la solución debía buscarse: o bien en la educación de los propios usuarios para que etiquetaran convenientemente sus aportaciones al piélago de información (aplicando por ejemplo los metadatos recomendados por el *Dublin Core*), o bien en la construcción y utilización de una *gran maquinaria* capaz de colocar cada documento en su lugar (Web Semántica).

Sin embargo, el método de Google fue el que se impuso y las razones pueden resumirse en dos puntos lógicamente independientes pero cuya combinación resultó explosiva:

- La búsqueda *intencional*: los usuarios saben lo que están buscando, acceden a la red con una intención.
- La búsqueda *social*: la tecnología *PageRank* consiste en priorizar en los listados que muestra Google aquellas páginas que son más enlazadas, de modo que poner un enlace a una página significa emitir un voto a favor de ella. Así, una página que haya sido muy enlazada (muy votada) adquiere mayor importancia y, además, los enlaces (votos) que ella misma emita tendrán más peso que el de otras menos enlazadas (votadas). Es decir, Google supone que los contenidos de aquellas páginas más votadas son las que la red colectiva de usuarios que las han visitado considera más relevantes.

De las taxonomías a las folksonomías: el caso de Eureka Swicki

Como se ha dicho, Google destacó respecto del resto de portales porque ofrecía una manera directa de acceder a los contenidos: ofrecía una página en blanco en cuyo centro aparecían poco más que una caja en la que meter la cadena de palabras buscada, un botón de “Buscar” y el logo de la compañía. Nada de carpetas, nada de navegación asistida. La tarea consistía en meter el término intencional requerido y esperar la respuesta. Obviamente, la manera en que estaba organizada la información debía tener alguna significación en los resultados mostrados en pantalla, y los creadores del algoritmo de *PageRank* supieron ver en los *hiperenlaces* el equivalente a las citas bibliográficas tradicionales de los artículos científicos.

Por consiguiente, si las respuestas sobre lo que uno busca dependen ciertamente de lo que otros previamente votaron como relevante, entonces se genera una fuerza en la que los estudios de orden social demandan atención. Como se sabe, uno de los puntos focales de la sociología de la ciencia se encuentra en la dinámica social de publicaciones, la tensión que éstas provocan sobre las relaciones que un científico tiene con respecto al resto de miembros de la comunidad científica vinculados a su especialidad o con respecto a las instituciones de las que espera financiación. En el nuevo entorno digital en el que los textos científicos y no científicos comienzan a

publicarse, cada vez más sin una copia gemela impresa, las cuestiones que tienen que ver con las relaciones sociales y las comunidades de interés mutuo están emergiendo con más fuerza.

La propia Google señala que sus editores son todos los usuarios de Internet de manera que se cumple lo que Ramesh Jain demandaba, a saber, “poner al usuario al volante”. Parece pues, por lo dicho hasta aquí, que la clave de los futuros buscadores estará en poner al usuario al volante con el propósito de que ajuste lo más posible su búsqueda a sus intenciones. Sin embargo, lo que Google ofrece actualmente a cualquier conductor es que si su intento es fallido tiene que empezar de nuevo poniendo a cero nuevamente el contador. Es decir, no hay continuidad en las búsquedas de manera que si el primer intento resultara fallido el buscador sacara partido de la situación resultante, esto es, reconociera la intención del usuario y afinara la búsqueda posterior.

De todos modos, el método de Google ha resultado hasta el día de hoy ser el más eficiente a la hora de buscar información y sus servicios especializados como Google Scholar o Google BookSearch sin duda afinarán mucho con los resultados. Realmente, la Google actual es una selva en la que todo el mundo opina acerca de los contenidos por lo que, aunque el método funciona mejor que el de sus rivales, podemos suponer que en muchos casos lo que valga para un lector no será relevante para otro. Sin embargo, han surgido en los últimos un tipo de iniciativas conocido como *folksonomías*, que incluye el término “folk” que proviene del inglés y cuya traducción al castellano es “gente”, en contraposición con las tradicionales *taxonomías*.

Las folksonomías se rigen por métodos similares al *PageRank* de Google, es decir, que confían en la capacidad de los lectores a la hora de discriminar y votar contenidos, pero dentro de comunidades específicas. Podría tratarse de grandes comunidades como la de los científicos en general, o bien podría referirse a comunidades de proporciones más modestas como por ejemplo la comunidad de especialistas en topología algebraica o la de los expertos en historia de la arqueología cartaginesa. La iniciativa Eureka Swicki es un buen ejemplo para esbozar la idea: uno puede crear su propia comunidad, a la que otros pueden apuntarse; una vez que el investigador escoge la comunidad (o las comunidades) que se ajusta a sus expectativas e inicia la búsqueda, los resultados que obtiene son aquellos que han pasado el primer filtro de Google y el segundo filtro de los miembros de su comunidad. Brevemente, cada comunidad Swicki altera los resultados ofrecidos por Google en función de los intereses de cada comunidad. Así, si el aficionado a la arqueología cartaginesa estuviera interesado en leer artículos que mencionaran en qué lugares de España se han hallado restos de Aníbal Barca, obtendría unos resultados más fiables filtrando a través de su comunidad que si lo hubiera hecho directamente con Google. Él mismo podría leer los diez primeros resultados obtenidos y aportar su granito de arena a la comunidad votando a algunas/os de las páginas o artículos de manera que contribuiría a aumentar el nivel de relevancia de los mismos y, por consiguiente, a que dichos textos aparezcan más arriba para el resto de miembros de su comunidad.

Buscando nuevos descriptores internos

Las redes sociales tipo folksonomías parecen prometedoras, pero no debemos olvidar que las citas que se realizan actualmente mediante hipervínculos son excesivamente

simples y que su complejidad puede incrementarse en el futuro. La propuesta consiste en asignar metadatos o etiquetas a los propios hipervínculos de modo que las citas sean multidimensionales. Los hipervínculos podrían no ser sólo asertivos, sino que podríamos en un futuro crear hipervínculos que indiquen una refutación o simplemente que la etiqueta del hipervínculo no apoye el contenido de otra a página u otro documento pero sí su diseño. En este segundo caso, sabríamos que el contenido no es fiable pero eso no nos impediría contemplar, por ejemplo, un maravilloso diseño gráfico.

En el modelo popperiano que proponemos, en el que abundan las conjeturas, los refrendos, las refutaciones y otros tipos de asertos imaginables, las citas pueden ser de muchos tipos. Un académico puede querer citar (mediante un hipervínculo) un artículo para apoyar una de sus tesis, o puede querer citarlo como ejemplo de lo que no hay que hacer, o quizá simplemente lo cita para aportar a su lector información adicional en la que éste pueda profundizar, esto es, una información perteneciente a un campo en el que él se reconoce menos competente pero que de buena fuente sabe que es fiable.

Todos estos descriptores internos a las relaciones que bosquejan los hipervínculos o los documentos mismos han de ayudar a los bibliotecarios a dar un mejor servicio en el entorno digital, a ofrecer a sus usuarios la información relevante en el tiempo más corto posible. En este panorama, los bibliotecarios han de ser más que nunca los arquitectos informacionales de la Red, aquellos que muestren los mejores itinerarios entre los documentos que pueblan la selva popperiana y que ofrezcan asesoría y consultoría a los investigadores a la hora de elegir la mejor comunidad o las mejores etiquetas para sus propósitos. Con la flexibilidad que ofrece el esqueleto digital, que aquí hemos descrito someramente pero que ofrece un terreno abonado para concretar infinidad de ideas, los usuarios tendrán en sus manos la mejor herramienta para moverse a lo largo y ancho de la biblioteca digital universal a base de muy pocos “clicks”. La bibliotecología tiene mucho terreno por delante donde innovar.

Bibliografía

Adair, W. C. (1955): “Citation Indexes for Scientific Literature”, *American Documentation* 6, pp. 31-32.

Garfield, Eugene (1957): “Breaking the Subject Index Barrier: a Citation Index for Chemical Patents”, *Journal of the Patent Office Society*, Vol. XXXIX, N° 8, August, pp. 583-595.

Garfield, Eugene (1964): “‘Science Citation Index’: a New Dimension in Indexing”, *Science*, Vol. 144, N° 3619, May 8, pp. 649-654.

Garfield, Eugene (1972): “The New Social Sciences Citation Index (SSCI) Will Add a New Dimension to Research on Man and Society”, *Essays of an Information Scientist*, Vol. 1, May 31, pp. 317-319.